ELSEVIER International Journal of Approximate Reasoning 25 (2000) 71–87

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Nonlinear nonnegative multiregressions based on Choquet integrals

Zhenyuan Wang [1], Kwong-Sak Leung, Man-Leung Wong, Jian Fang [2], Kebin Xu [*,3]

Department of Computer Science & Engineering, Chinese University of Hong Kong, Shatin, NT, Hong Kong

Received 1 May 1998; accepted 1 June 2000

Abstract

Using a nonadditive set function to describe the interaction among attributes, a new nonlinear nonnegative multiregression is established based on Choquet integrals with respect to the set function. Regarding the values of the set function as unknown regression parameters, an evolutionary computation can be used to determine them when necessary data are available. Such a model is a generalization of the traditional linear multiregression. It provides an effective regression tool in some real problems where the linear multiregression model and the second-order multiregression model fail. This new method has a wide applicability in the areas of information fusion and data mining, as well as in the areas of decision making, image processing, pattern recognition, medical and industrial diagnoses, and expert systems. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Nonlinear regressions; Information fusion; Data mining; Nonnegative monotone set functions; Choquet integrals; Optimization; Evolutionary algorithms

* Corresponding author.

E-mail address: kbxu@cse.cuhk.edu.hk (K. Xu).

[1] On leave from the Department of Mathematics, Hebei University, Baoding, Hebei 071002, P.R. China.

[2] On leave from the Department of Information and Control Engineering, Shanghai Jiaotong University, Shanghai 200030, P.R. China.

[3] On leave from the Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, P.R. China.

PII: S0888-613X(00)00048-7

1. Introduction

In information fusion and data mining, the most common aggregation tools are the weighted average method and the linear regression. They are essentially the Lebesgue-like integrals on discrete spaces. These linear methods need a basic supposition that there is no interaction among attributes and, therefore, the global contribution of a set of attributes to the target is just the simple sum of contributions from each attribute in the set. However, in many real problems, the inherent interaction cannot be ignored. In this case, a nonlinear model should be used. Though there are many different types of traditional nonlinear models, such as second-order models, exponential or logarithmic transformations [3,8,9], that can be chosen, due to their lack of inherent supporting mechanisms, these models also fail in most real problems.

Recently, the introduction of concepts of nonadditive set function (also called fuzzy measure) and relevant nonlinear integral [1,10,12,16] ushered in a new way to deal with the above problems of data analysis. Based on the concepts of nonadditive set function and relevant nonlinear integral, this paper provides a new nonlinear multiregression model that can be widely applied in information fusion [5,6,22,24] and data mining, as well as in the areas of decision making, image processing, pattern recognition, medical and industrial diagnoses, and expert systems.

In Section 2, an overview of the traditional multiregression model is given. Section 3 provides the fundamental knowledge of nonadditive set functions and relevant nonlinear integrals. Based on this knowledge, a new nonlinear multiregression model is established in Section 4. To determine the regression coefficients that are the values of the nonadditive set function used in the model, Section 5 presents an adaptive evolutionary algorithm with a flow chart. A successful test example is shown in Section 6. Finally, several conclusions are given in Section 7.

2. Traditional multiregressions

One of the traditional methods used in data analysis is the linear multiregression. It has a form as follows:

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n + N(a_0, \sigma^2),$$

where Y is the dependent variable, $X_1, X_2, \ldots, X_n$ are n independent variables, and $N(a_0, \sigma^2)$ is a normal distributed random variable with mean a_0 and variance σ^2. It also can be viewed as an application of the Lebesgue-like integral [4] in the multisource fusion of numerical information. Indeed, let us

assume that $a_j \geqslant 0$ for all j and take $X = \{x_1, x_2, \ldots, x_n\}$, where x_j is the jth information source and the numerical information obtained from x_j is X_j, $j = 1, 2, \ldots, n$. If we define a set function μ, which indicates the importance of each x_j, on the class of all singletons $\{x_j\}$ by $\mu(\{x_j\}) = a_j,\ j = 1, 2, \ldots, n$, then μ can be uniquely extended to be an additive measure onto $\mathscr{P}(X)$, the power set of X. Thus, the Lebesgue-like integral $\int f \, \mathrm{d}\mu$ can be used as a comprehensive numerical assessment for the target, where the integrand f is a function defined on X whose value at x_j is $f(x_j) = X_j,\ j = 1, 2, \ldots, n$. Hence, the above-mentioned linear multiregression model can be expressed as

$$Y = \int f \, \mathrm{d}\mu + N(a_0, \sigma^2)$$

and can be employed to determine the additive measure μ from the data of $(Y, f(x_1), f(x_2), \ldots, f(x_n))$.

However, due to inherent interactions, which are essentially different from a correlationship, to the target among diverse information sources, such a linear model becomes invalid in many real problems. The use of some improved nonlinear multiregression models, which are based on an artificially chosen nonlinear function of a given multivariate, is also very limited due to the lack of a justification for such a choice from a great variety of nonlinear functions. Usually, just for convenience, quadratic or more general second-order functions are used [3,8,9]. For instance, Mendenhall and Sincich [8, pp. 716–719] used a complete (or full in some books) second-order model in an example with two independent variables:

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_1 X_2 + a_4 X_1^2 + a_5 X_2^2 + N(a_0, \sigma^2).$$

Though the result for this example seemingly shows that the model is satisfactory (passing the F test and the coefficient of determination R^2 being as high as 0.993), some people will doubt the reality of the data. In the literature, mostly, there are no justifications given for the choice of a second-order function. Such a choice is only for the purpose of convenience in the case of failure in using a linear model since the second-order function is the simplest nonlinear function. Now, a question arises naturally: In some real problems, if the F test cannot be passed and/or the coefficient of determination R^2 is not large enough for a second-order model, what should we do? To deal with these real problems, this paper provides a reasonable means in which nonadditive set functions are used to replace the additive measure and, relevantly, instead of the Lebesgue-like integral, the Choquet integral [1,2,10,17] or any other suitable nonlinear integral (for example, the Sugeno integral [11,15,16], the pan-integral [16,18,23], the natural extension [12,17,19], and Wang's integral [21,25]) is used. The details are given in the following sections.

3. Nonadditive set functions and Choquet integrals

Let $\mathscr{F}$ be a nonempty class of subsets of X with $\varnothing \in \mathscr{F}$ and $X \in \mathscr{F}$. A nonnegative monotone set function, μ, is a mapping from $\mathscr{F}$ to $[0,\infty)$ satisfying the following conditions:

$$\mu(\varnothing) = 0, \tag{1}$$

$$A \subset B \Rightarrow \mu(A) \leqslant \mu(B) \quad \forall A, B \in \mathscr{F}. \tag{2}$$

If $\mu(X) = 1$, then μ is called regular. Since, in current research, space X is always finite, the continuity requirement on μ is insignificant and, therefore, such a set function is also called a fuzzy measure [11,15,16]. It is convenient to take the power set of X, $\mathscr{P}(X)$, as $\mathscr{F}$ when X is finite. μ is nonadditive in general. The nonadditivity of μ means that, to the comprehensive numerical assessment for the target, the combinative contribution of information sources in a set may be more or less than the sum of the contributions of each source in the set, and it describes the inherent interaction among the sources. The value of μ at a set of sources indicates the possible maximum rate of contribution, including the interaction, to the comprehensive assessment for the target due to the sources in this set, and we can understand it as the relative importance of this source set regarded as a whole. So, we also call set function μ an importance measure. When we use a nonnegative monotone set function to replace the additive measure for information fusion, the Lebesgue-like integral is unavailable generally. We must use some nonlinear integrals with respect to the nonadditive set function. The Choquet integral is one of them that is defined as follows:

$$\int f \, \mathrm{d}\mu = \int_0^\infty \mu(F_\alpha) \, \mathrm{d}\alpha,$$

where f is a nonnegative measurable function on $(X, \mathscr{F})$ in the sense that $\{x \mid f(x) \geqslant \alpha, x \in X\} \in \mathscr{F}$ for any $\alpha \in [0,\infty)$, and $F_\alpha = \{x \mid f(x) \geqslant \alpha, x \in X\}$ is the α-cut set of f. Since the right-hand side is a Riemann integral, and its integrand, $\mu(F_\alpha)$, is a nonincreasing function of α, the above integral is well-defined. In most real problems, X is finite. In this case, say, $X = \{x_1, x_2, \ldots, x_n\}$, the Choquet integral is also well-defined with respect to a nonnegative set function that is not necessarily monotone. To calculate its value, first, the values of f, $\{f(x_1), f(x_2), \ldots, f(x_n)\}$, should be rearranged in a nondecreasing order, that is,

$$f(x_1^*) \leqslant f(x_2^*) \leqslant \cdots \leqslant f(x_n^*),$$

where $(x_1^*, x_2^*, \ldots, x_n^*)$ is a permutation of $(x_1, x_2, \ldots, x_n)$. Then, the value of the Choquet integral can be obtained by expression

$$\int f \, \mathrm{d}\mu = \sum_{i=1}^{n} \left[f(x_i^*) - f(x_{i-1}^*) \right] \mu\left(\left\{ x_i^*, x_{i+1}^*, \ldots, x_n^* \right\} \right),$$

where $f(x_0^*) = 0$.

The following is an example of using nonnegative monotone set function and the Choquet integral as an aggregation tool in information fusion.

Example 1. To detect an object that may be a tank or an armored personnel carrier (APC), three different means, x_1, x_2, and x_3, are adopted, and they are regarded as information sources, that is, $X = \{x_1, x_2, x_3\}$. Suppose that the importance degrees (the values of nonnegative monotone set function μ) of these means and their combinations are given as follows:

$\mu(\varnothing)$	0
$\mu(\{x_1\})$	0.20
$\mu(\{x_2\})$	0.10
$\mu(\{x_3\})$	0.40
$\mu(\{x_1, x_2\})$	0.34
$\mu(\{x_1, x_3\})$	0.75
$\mu(\{x_2, x_3\})$	0.58
$\mu(\{x_1, x_2, x_3\})$	1

Also suppose that, through means x_1, x_2, and x_3, the marginal evaluations (information) for tank hypothesis, denoted by f_1, f_2, and f_3, respectively, are obtained:

$$f_1 = 0.85, \quad f_2 = 0.21, \quad f_3 = 0.90.$$

The Choquet integral with respect to μ can be used to obtain a comprehensive assessment for the tank hypothesis. Its value is calculated as follows:

Let

$$x_1^* = x_2, \quad x_2^* = x_1, \quad x_3^* = x_3.$$

Then

$$f(x_1^*) = 0.21, \quad f(x_2^*) = 0.85, \quad f(x_3^*) = 0.90,$$

and

$$\begin{aligned} \hat{E} = \int f \, \mathrm{d}\mu &= \sum_{j=1}^{3} \left[f(x_j^*) - f(x_{j-1}^*) \right] \mu\left(\left\{ x_i^*, x_{i+1}^*, \ldots, x_3^* \right\} \right) \\ &= 0.21 \times 1 + 0.64 \times 0.75 + 0.05 \times 0.40 \\ &= 0.71. \end{aligned}$$

Number $\hat{E} = 0.71$ is the comprehensive assessment for tank hypothesis by using these three means.

There are some examples to show that we have to use nonadditive set functions and relevant nonlinear integrals to replace classical additive measures and the traditional integrals in some decision making, evaluation, and other real problems [16, Example 9.11].

The Choquet integral has a very intuitive explanation where nonnegative set function μ is interpreted as an efficiency function. This is shown in the following example as well as in [10].

Example 2. There are three workers a, b, and c working for $f(a) = 10$, $f(b) = 15$, and $f(c) = 7$ days, respectively, to manufacture a kind of product. Without any manager, they begin to work from the same day. Their efficiencies of working alone are 5, 6, and 8 products per day, respectively. Their joint efficiencies are not the simple sum of the corresponding efficiencies given above, but are listed as follows:

$\{a, b\}$	14
$\{a, c\}$	7
$\{b, c\}$	16
$\{a, b, c\}$	18

These efficiencies can be regarded as a nonnegative set function (not necessarily monotone), μ, defined on the power set of $X = \{a, b, c\}$ with $\mu(\varnothing) = 0$ (the meaning is that there is no product if there is no worker). Here inequality $\mu(\{a, b\}) > \mu(\{a\}) + \mu(\{b\})$ means that a and b have a good cooperation, while inequality $\mu(\{a, c\}) < \mu(\{a\}) + \mu(\{c\})$, and even $\mu(\{a, c\}) < \mu(\{c\})$, means that a and c have a very bad relationship and they are not suitable for working together. Set function μ is not necessarily monotone and is called an efficiency measure. In such a simple manner, during the first 7 days, all workers work together with efficiency $\mu(\{a, b, c\})$, and the number of products is $f(c) \cdot \mu(\{a, b, c\}) = 7 \times 18 = 126$; during the next $f(a) - f(c)$ days, workers a and b work together with efficiency $\mu(\{a, b\})$, and the number of products is $[f(a) - f(c)]\mu(\{a, b\}) = 3 \times 14 = 42$; during the last $f(b) - f(a)$ days, only b works with efficiency $\mu(\{b\})$, and the number of products is $[f(b) - f(a)]\mu(\{b\}) = 5 \times 6 = 30$. Thus, the value of (c)$\int f \, d\mu$, 198, is just the total number of products manufactured by these workers during these days.

The nonnegative set function above is a generalization of the classical additive measure while the Choquet integral is a generalization of the Lebesgue-like integral. In a special case that the nonnegative set function is additive, the Choquet integral coincides with the Lebesgue like integral.

4. A new model of nonlinear nonnegative multiregression

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of attributes, monotone set function $\mu: \mathscr{P}(X) \to [0, 1]$ satisfying $\mu(\varnothing) = 0$ and $\mu(X) = 1$ be an importance measure.

Set function μ describes the importance of all individual attributes as well as all combinations of attributes in X. The marginal evaluation from all individual attributes is denoted by a function, $f : X \to [0, \infty)$, that is, $f(x_i)$ is the evaluation (or, the information received) from attribute x_i, $i = 1, 2, \ldots, n$. Suppose that a variable, Y, depends on the attributes in the following way:

$$Y = q \int f \, \mathrm{d}\mu + N(c, \sigma^2),$$

where the integral may be chosen from the Choquet integral, the natural extension (a nonlinear functional introduced in the theory of imprecise probabilities), and some type of nonlinear integrals else, q is a proportional divisor, and $N(c, \sigma^2)$ is a normally distributed random variable with mean c and variance σ^2. This relation can be regarded as a multi-input single-output system with a random interference, where the input is function f while the output is Y (see Fig. 1). After fixing the type of integral, such a system is depicted by set function μ and constants q, c, and σ. By observing the input and output of the system for l times independently, the input–output data with size l may be obtained as follows:

$$\begin{array}{cccc} f_{11} & f_{12} & \cdots & f_{1n} \quad Y_1 \\ f_{21} & f_{22} & \cdots & f_{2n} \quad Y_2 \\ \vdots & & & \\ f_{l1} & f_{l2} & \cdots & f_{ln} \quad Y_l \end{array}$$

The least-square method can be used to determine the unknown parameters of the system from the data. These parameters are the values of μ and constants q, c, and σ. Indeed, μ, q, and c can be determined according to the criteria that

$$e = \sqrt{\frac{1}{l} \sum_{j=1}^{l} \left(Y_j - q \int f_j \, \mathrm{d}\mu - c \right)^2}$$

is minimized, and σ can be estimated in terms of e, where f_j is a function on X with $f_j(x_i) = f_{ji}$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, l$.

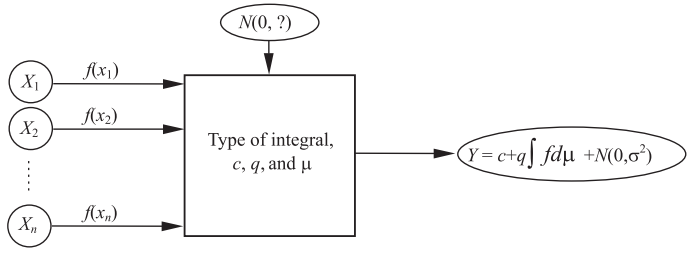

Fig. 1. Regarding the nonlinear multiregression as a multi-input single-output system.

As mentioned in Section 3, when μ is additive, the Choquet integral coincides with the Lebesgue-like integral. Therefore, the above model degenerates into the classical linear multiregression in this case. Generally, the model shown here is a nonlinear multiregression. It is a generalization of the traditional linear multiregression and there is a natural mechanism as a justification for using such a new nonlinear multiregression model. It is hopeful that our new model will lead to ideal results in some real problems where the traditional linear model and the quadratic model do not work well. It is expected, by using this model in real problems, that the precision and validity of a forecast or an inference will be raised. Hence, this new regression model has a wide applicability.

The crux of using this model is to determine the unknown regression parameters, i.e., values of the nonnegative set function and constants c and q. Unlike the traditional linear multiregression for which the regression coefficients can be obtained by solving a system of linear equations, there is no classical mathematical method which can be used to determine the unknown parameters in our new regression model. Fortunately, the determination can be done through a new computational method so-called soft computing, for example, training appropriate neural networks [13] or using special genetic algorithms [14,20,24], when sample data are available. In this paper, an evolutionary algorithm is shown in the next section.

5. An adaptive evolutionary algorithm

To determine the unknown parameters in the above-mentioned nonlinear multiregression model from the given data, the logic flow of our adoptive evolutionary algorithm with double optimization is summarized in Fig. 2.

(1) Input given data consisting of n, l, (f_{ji}), and (Y_j) in the following form, where n is the number of considered attributes, l is the size of data,

$$(f_{ji}) = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & & & \\ f_{l1} & f_{l2} & \cdots & f_{ln} \end{pmatrix},$$

and

$$(Y_j) = (Y_1 \quad Y_2 \quad \cdots \quad Y_l).$$

This step includes an examination of the completeness for the input data. In this step, users can choose gene's bit length m according to the required precision for the result and preferred population size S. The defaults of m and S

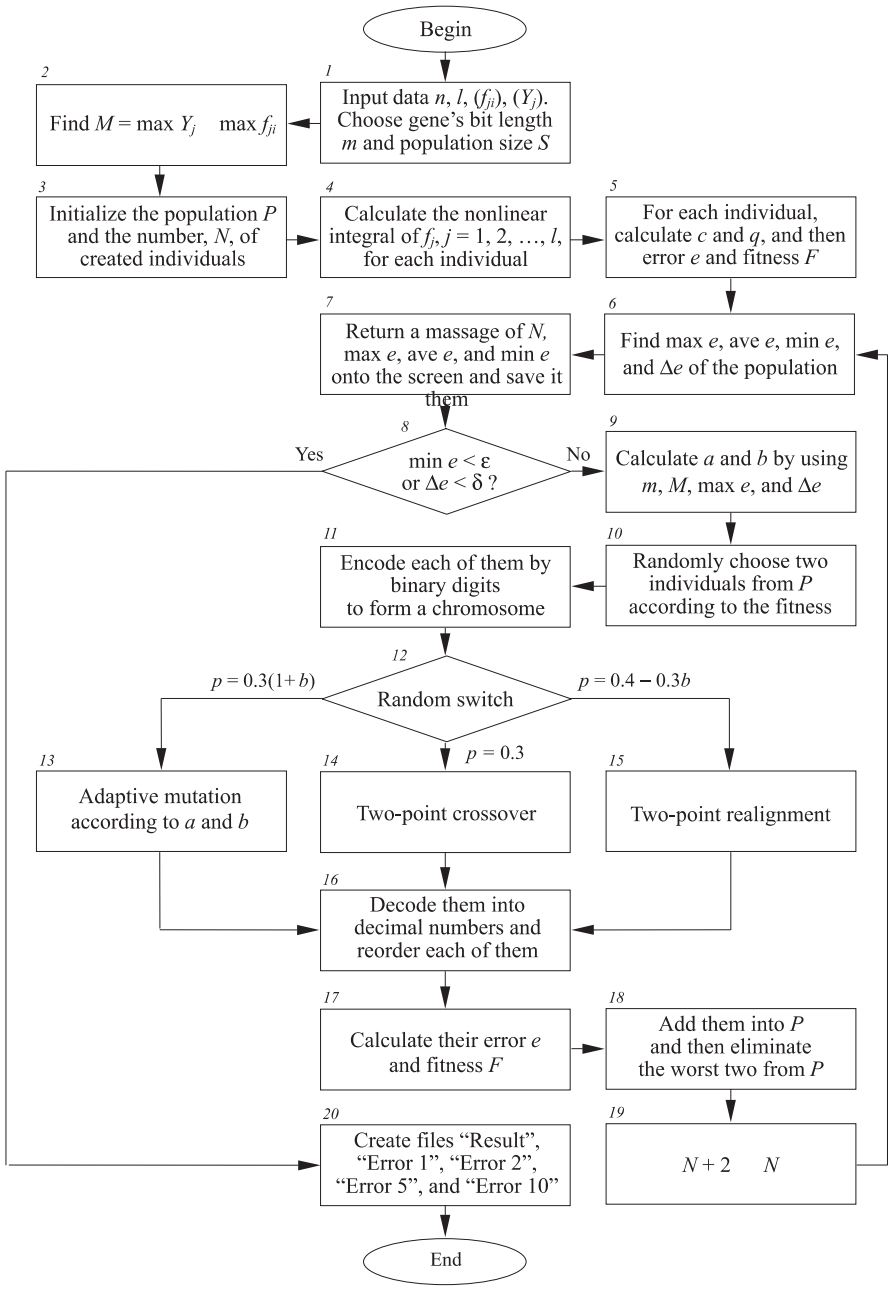

Fig. 2. Flow chart of an evolutionary algorithm used for nonlinear nonnegative multiregressions.

are 10 (the corresponding precision is three decimal digits) and 100, respectively. Users can also choose the type of nonlinear integrals: the Choquet integral or some other type of nonlinear integrals.

(2) Find M, which is used to measure the relative size of error e for each individual, defined as

$$M = \max_{1 \leqslant j \leqslant l} Y_j \vee \max_{\substack{1 \leqslant j \leqslant l \\ 1 \leqslant i \leqslant n}} f_{ji}.$$

(3) Initialize population P that consists of S individuals. Each individual in the population is a nonnegative monotone set function, μ, satisfying conditions $\mu(\varnothing) = 0$ and $\mu(X) = 1$. It consists of $2^n - 2$ real-valued variables in $[0, 1]$. In the power set $\mathscr{P}(X)$, sets $E_k,\ k = 0, 1, 2, \ldots, 2^n - 1$, are arranged in the following way: binary number $(k)_2 = b_n^{(k)} b_{n-1}^{(k)}, \ldots, b_1^{(k)}$ is determined by

$$b_i^{(k)} = \begin{cases} 1 & x_i \in E_k, \\ 0 & x_i \notin E_k, \end{cases} \quad i = 1, 2, \ldots, n.$$

That is, $E_0 = \varnothing,\ E_1 = \{x_1\}, E_2 = \{x_2\}, E_3 = \{x_1, x_2\}, E_4 = \{x_3\}, E_5 = \{x_1, x_3\}, E_6 = \{x_2, x_3\},\ E_7 = \{x_1, x_2, x_3\},\ E_8 = \{x_4\},\ E_9 = \{x_1, x_4\},\ E_{10} = \{x_2, x_4\}, \ldots$ To form an individual, use a random number generator to create $2^n - 2$ random numbers in $[0, 1]$ as the values of μ at $E_1, E_2, \ldots, E_{2^n-2}$, respectively. Then, use the following reorder algorithm to exchange these values such that the monotonicity is held for μ.

Reorder algorithm. For each E_k, from $k = 2$ to $k = 2^n - 2$, check each E_h, from $h = 1$ to $h = k - 1$, whether $E_h \subset E_k$ with $\mu(E_h) > \mu(E_k)$. If yes, exchange their values.

After S individuals being created, let $N = S$.

(4) For each $f_j,\ j = 1, 2, \ldots, l$, calculate the value of the chosen nonlinear integral with respect to each importance measure in the population.

(5) For each individual μ (an importance measure) in population P, use the data to calculate its error e by expression

$$e = \sqrt{\frac{1}{l} \sum_{j=1}^{l} \left(Y_j - q \int f_j \, \mathrm{d}\mu - c \right)^2},$$

where

$$q = \frac{\sum_{j=1}^{l} \left(\int f_j \, \mathrm{d}\mu - \overline{I} \right) \left(Y_j - \overline{Y} \right)}{\sum_{j=1}^{l} \left(\int f_j \, \mathrm{d}\mu - \overline{I} \right)^2},$$

$$c = \overline{Y} - q\overline{I},$$

while

$$\overline{Y} = \frac{1}{l}\sum_{j=1}^{l} Y_j$$

and

$$\overline{I} = \frac{1}{l}\sum_{j=1}^{l} \int f_j \, \mathrm{d}\mu.$$

The corresponding fitness is defined as

$$F = \frac{1}{1 + (5e/M)} \quad \text{(weak form)}$$

or

$$F = 2^{-(5e/M)} \quad \text{(strong form)}.$$

After running the relevant programs, it is possible to compare these two forms to show which is better.

(6) Find the maximum error and the minimum error, denoted by max e and min e, respectively, in the population. Calculate the average error, ave e, of the population and $\Delta e = \max e - \min e$.

(7) Return a message of N, max e, ave e, and min e onto the screen. Save it in a file entitled "Error".

(8) Check whether $\min e < \delta = 2^{-2m}$ or $\Delta e < \varepsilon = 2^{-3m}$. If yes, go to (20); if no, go to the next step.

(9) Calculate

$$a = \frac{-\log_2(\max e/M)}{m} \wedge 1$$

and

$$b = \frac{-\log_2(\Delta e/M)}{m} \wedge 1.$$

These two parameters will be used to adjust probability distributions that control active operators in the evolutionary process adaptively.

(10) Randomly choose two individuals from the population as parents. The probability to pick an individual out is proportional to its fitness. The chance of picking two identical individuals is very rare.

Encode the values of each individual picked out above in binary digits to form a chromosome in the following way: for each $\mu(E_j)$, take the m-bit binary representation of $\lfloor 2^m \times \mu(E_j) \rfloor$ as a gene, then align all these genes according to the order of sets E_j, $j = 1, 2, \ldots, 2^n - 2$, where symbol $\lfloor r \rfloor$ denotes the integer

part of nonnegative real number r. Hence, the bit length of each chromosome is $m \times (2^n - 2)$.

(12) Among three given operators, adaptive mutation, two-point crossover, and two-point realignment, use a random switch to determine which one will be adopted to reproduce offspring from the chosen parents. The random switch is constituted by a uniformly distributed random variable ξ in $[0, 1]$ and a partition of $[0, 1]$ with length $0.3(1 + b)$, 0.3, and $0.4 - 0.3b$. If $\xi \in [0, 0.3(1 + b)]$, then go to (13), and an adaptive mutation will be adopted. If $\xi \in (0.3(1 + b), 0.3(2 + b)]$, then go to (14), and a two-point crossover will be adopted. If $\xi \in (0.3(2 + b), 1]$, then go to (15), and a two-point realignment will be adopted. In such a way, the probability to choose mutation will be increased when Δe becomes small. The converse situation appears for two-point realignment.

(13) An adaptive k-point mutation is adopted here. Number k is a random positive integer between 1 and n determined by a random number γ in $[0, 1]$ with expression $k = \lceil n\gamma \rceil \vee 1$, where symbol $\lceil r \rceil$ denotes the smallest integer that is not smaller than the nonnegative real number r. The bit location of each of these k points, at which the binary digit is switched from 0 to 1 or from 1 to 0, is independently determined by a random number, ζ, in the following way: the bit location from the left of the chromosome is computed by formula

$$\lfloor \zeta \times (2^n - 2) \rfloor \times m + \lceil (\theta^{-1}(d(\zeta \times (2^n - 2)))) \times m \rceil \vee 1,$$

where d is the symbol of decimal part and function θ^{-1} is defined by

$$\theta^{-1}(s) = \begin{cases} (a(1 - a + b)s)^{1/2}, & s \in [0, a/(1 - a + b)], \\ (1/2)((1 - a + b)s + a), & s \in (a/(1 - a + b), (2b - a)/(1 - a + b)], \\ 1 - ((1 - s)(1 - a + b)(1 - b))^{1/2}, & s \in ((2b - a)/(1 - a + b), 1]. \end{cases}$$

Through transformation θ^{-1}, a uniform distribution is remoulded to be a biased distribution that mostly focuses on lower bits in each gene when error e or Δe is small while on higher bits when Δe is large. This is why the mutation used in the algorithm is said to be adaptive. By using such a transformation, it is expected that the convergence speed of the algorithm will be raised. After taking the mutation operator on each of two chromosomes formed in (11), go to (16).

(14) Two random numbers are used to determine two points at which both chromosomes formed in (11) are cut into three segments. Then exchange their middle segments to form two new chromosomes. After this, go to (16).

(15) For each chromosome obtained in (11), independently use two random numbers to determine two points at which the chromosome is cut into three segments, and then using random switches to choose one from 48 different types (with different orders or different directions of segments) to realign these three segments to be a new chromosome.

(16) Decode two new chromosomes obtained above into decimal numbers and use the algorithm given in (3) to reorder them such that each of them represents an importance measure.

(17) Calculate error e and fitness F of new importance measures by using the formulas given in (5).

(18) Add these two importance measures to the population. Then, according to the fitness, eliminate the worst two individuals from the population.

(19) Add 2 on number N to replace the original N. Now N is the number of importance measures created so far. Then, return to (6).

Before the algorithm stops, create a file called "Result" in which the current values of N, $\min e$, corresponding constants q and c, importance measure μ, and $Y_j^* = q\int f_j \,\mathrm{d}\mu + c, j = 1, 2, \ldots, l$, are listed. The values of μ are listed in the following form:

Set	Value of μ
$\varnothing$	0
$\{x_1\}$	⋮
$\{x_2\}$	
$\{x_1, x_2\}$	
$\{x_3\}$	
$\{x_1, x_3\}$	
$\{x_2, x_3\}$	
$\{x_1, x_2, x_3\}$	
$\{x_4\}$	
$\{x_1, x_4\}$	
$\{x_2, x_4\}$	
⋮	
$\{x_2, x_3, \ldots, x_n\}$	
X	1

File "Error 1" is formed by picking the tuples (N and the corresponding errors) from file Error for $N = S, 2S, 3S, \ldots$, until its end (including the last tuple even if the last N in the file is not a multiple of S). Similarly, we form files "Error 2", "Error 5", and "Error 10" where N is a multiple of $2S$, $5S$, and $10S$, respectively. These files are used to plot the curves showing the convergence rate of the algorithm as presented in Fig. 3.

6. Example

The algorithm is implemented in program language C++ as well as in Visual C++. It has been run for a number of examples successfully. The following is one of them where $m = 10$, $S = 100$, and the seed is 3.

Example 3. Let $X = \{x_1, x_2, x_3\}$. The data with size $l = 12$ are given as follows:

f_1	f_2	f_3	Y
0.3	0.9	1.4	8.075
1.5	1.2	0.7	8.725
1.3	1.8	0.1	8.675
0	0.1	0.8	6.375
0.6	0.9	0	7.275
1.4	0.5	0.7	8.125
1.7	1.8	1.1	9.775
0.2	0.9	0.4	7.525
1	1.3	0.5	8.375
1.1	0	1.2	7.950
0.8	0.9	0.6	7.925
0.3	0.7	1.8	7.825

After inputting the data and running the program, a result is obtained as follows:

$N = 8536,\ q = 2.481204585,\ c = 6.005856850,\ e = 0.007331408.$

Set	Value of μ
$\varnothing$	0
$\{x_1\}$	0.312500
$\{x_2\}$	0.495117
$\{x_1, x_2\}$	0.609375
$\{x_3\}$	0.092773
$\{x_1, x_3\}$	0.705078
$\{x_2, x_3\}$	0.820313
X	1

The convergence rate is demonstrated in Fig. 3. The error e is extremely small. In fact, there exists a precise solution in this example:

$q = 2.5,\ c = 6.$

Set	Value of μ
$\varnothing$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.5
$\{x_1, x_2\}$	0.6
$\{x_3\}$	0.1
$\{x_1, x_3\}$	0.7
$\{x_2, x_3\}$	0.8
X	1

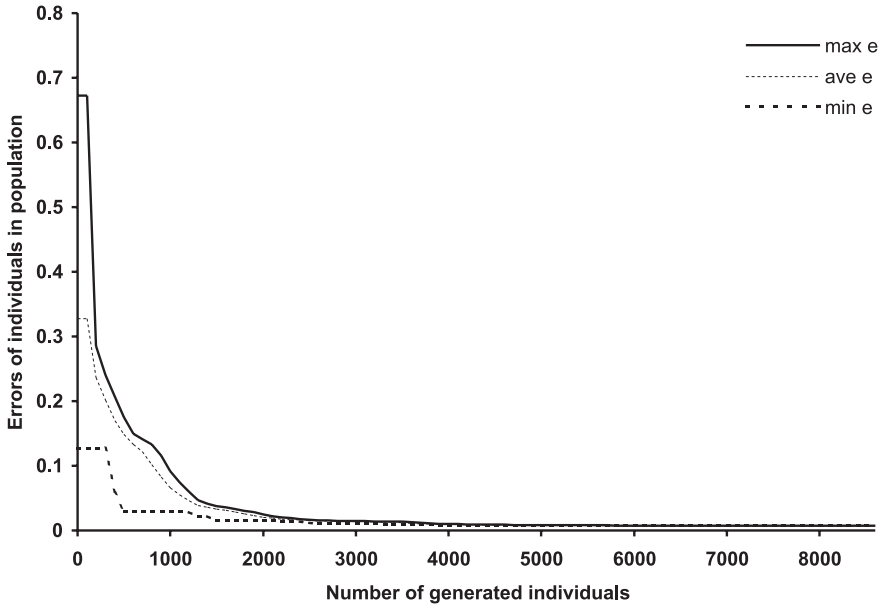

Fig. 3. Convergence rate in Example 3.

We can see that the values of μ in the result and in the precise solution are very close.

7. Conclusions

The new nonlinear model shown in this paper generalizes the traditional linear multiregression model. There are intuitive inherent mechanisms that support our model to represent some nonlinear multivariate relations in many real problems properly.

The convergence speed is acceptable. When the population size is 100, for most problems involving three attributes and one dependent variable, the running time on Pentium II is just a few seconds. But the running time of the program will exponentially increase as the number of attributes go up. The precision of the solution is also satisfactory. Usually, the error appears only after the second decimal place. This is enough for decision making. The precision will be raised if a bigger population size is assigned.

To obtain a fine regression effect, the sample size is required to be at least 1.5 times of the number of unknown parameters, 2^n, where n is the number of attributes. A sample size of 3–5 times of 2^n is preferred.

Some generalizations of the current nonlinear multiregression model are in progress. It is also hopeful to establish a network construction for a given database by such a nonlinear multiregression model [7].

Acknowledgements

This research is partially supported by a Hong Kong Government RGC Earmarked Grant, Ref. No. CUHK 4161/97E and a Lignan University Direct Grant, Ref. No. RES-001/199.

References

[1] L.M. de Campos, M.J. Bolaños, Characterization and comparison of Sugeno and Choquet integrals, Fuzzy Sets and Systems 52 (1992) 61–67.
[2] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1954) 131–295.
[3] J.L. Devore, Probability and Statistics for Engineering and the Sciences, fourth ed., Duxbury Press, Belmont, CA, 1995.
[4] P.R. Halmos, Measure Theory, Van Nostrand, New York, 1967.
[5] J.M. Keller, H. Qiu, H. Tahani, Fuzzy integral and image segmentation, in: Proceedings of the NAFIPS'86, New Orleans, 1986, pp. 324–338.
[6] J.M. Keller, J. Osborn, Training the fuzzy integral, International Journal of Approximate Reasoning 15 (1996) 1–24.
[7] K.S. Leung, M.L. Wong, W. Lam, Z. Wang, Discovering nonlinear-integral networks from databases using evolutionary computation and minimum description length principle, Proc. IEEE SMC'98, 2354–2359.
[8] W. Mendenhall, T. Sincich, Statistics for Engineering and the Sciences, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[9] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, Wiley, New York, 1994.
[10] T. Murofushi, M. Sugeno, M. Machida, Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets and Systems 64 (1994) 73–86.
[11] M. Sugeno, Theory of Fuzzy integral and its applications, Ph.D. Dissertation, Tokyo Institute of Technology, 1974.
[12] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[13] J. Wang, Z. Wang, Using neural networks to determine Sugeno measures by statistics, Neural Networks 10 (1) (1997) 183–195.
[14] W. Wang, Z. Wang, G.J. Klir, Genetic algorithms for determining fuzzy measures from data, Journal of Intelligent & Fuzzy Systems 6 (2) (1998) 171–183.
[15] Z. Wang, The autocontinuity of set function and the fuzzy integral, Journal of Mathematical Analysis and applications 99 (1984) 195–218.
[16] Z. Wang, G.J. Klir, Fuzzy Measure Theory, Plenum Press, New York, 1992.
[17] Z. Wang, G.J. Klir, Choquet integrals and natural extensions of lower probabilities, International Journal of Approximate Reasoning 16 (1997) 137–147.
[18] Z. Wang, G.J. Klir, PFB-integrals and PFA-integrals with respect to monotone set functions, International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 5 (2) (1997) 163–175.

[19] Z. Wang, G.J. Klir, J. Swan-Stone, K. Xu, An algorithm for calculating natural extensions with respect to lower probabilities, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), Fuzzy Measures and Integral – Theory and Applications, Springer, Heidelberg, 2000, pp. 467–476.
[20] Z. Wang, K.S. Leung, J. Wang, Determining nonnegative monotone set functions based on Sugeno's integral: an application of genetic algorithms, Fuzzy Sets and Systems 112 (2000) 155–164.
[21] Z. Wang, K.S. Leung, M.L. Wong, J Fang, A new type of nonlinear integrals and the computation algorithm, Fuzzy Sets and Systems 112 (2000) 223–231.
[22] Z. Wang, K.S. Leung, J. Wang, Genetic algorithms used for determining nonadditive set functions in information fusion, in: Proceedings of the IFSA'97, vol. 1, 1997, pp. 518–521.
[23] Z. Wang, W. Wang, G.J. Klir, Pan-integrals with respect to imprecise probabilities, International Journal of General Systems 25 (1996) 229–243.
[24] Z. Wang, K. Xu, J. Wang, G.J. Klir, Using genetic algorithms to determine nonnegative monotone set functions for information fusion in environments with random perturbation, International Journal of Intelligent Systems 14 (1999) 949–962.
[25] K. Xu, Z. Wang, K.S. Leung, Using a new type of nonlinear integral for multi-regression: an application of evolutionary algorithms in data mining, Proc. IEEE SMC'98 (1998) 2326–2331.